*Citation for published version:*
Attridge, N, Keogh, E & Eccleston, C 2019, 'An investigation of the effect of experimental pain on logical reasoning', *Pain*, vol. 160, no. 5, pp. 1093-1102. https://doi.org/10.1097/j.pain.0000000000001490

*DOI:*
10.1097/j.pain.0000000000001490

*Publication date:*
2019

*Document Version*
Peer reviewed version

Link to publication

**University of Bath**

## Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

(in press, PAIN)

# An investigation of the effect of experimental pain on logical reasoning.

Nina Attridge[1]*, Edmund Keogh[2] and Christopher Eccleston[2,3]

[1]Mathematics Education Centre, School of Science, Loughborough University, UK

[2]Centre for Pain Research, University of Bath, UK

[3] Department of Clinical and Health Psychology, Ghent University, Belgium

*****Corresponding author. Address: 0.28 Schofield Building, School of Science, Loughborough University, LE11 3TU, UK. Tel: +441509 223315. Email: n.f.attridge@lboro.ac.uk. Website: www.lboro.ac.uk.

Figures: 2

Tables: 0

## Abstract

Pain disrupts attention in order to prioritise avoidance of harm and promote analgesic behaviour. This could in turn have negative effects on higher-level cognitions which rely on attention. In the current paper we examined the effect of thermal pain induction on three measures of reasoning: the Cognitive Reflection Test, Belief Bias Syllogisms task, and Conditional Inference task. In Experiment 1, the thermal pain was set at each participant's pain threshold. In Experiment 2, it was set to a minimum of 44°C or 7/10 on a VAS scale (whichever was higher). In Experiment 3, performance was compared in no pain, low intensity pain, and high intensity pain conditions. We predicted that the experience of pain would reduce correct responding on the reasoning tasks. However, this was not supported in any of the three studies. We discuss possible interpretations of our failure to reject the null hypothesis and the importance of publishing null results.



## Summary

Pain disrupts attention, which in turn is important for logical reasoning. However, across three experiments, induced pain did not affect participants' logical reasoning behaviour.



Keywords:

Pain; Cognition; Reasoning; Logic; Disruption; Executive function; Cognitive Intrusion.

1. Introduction

There is evidence that pain disrupts attention [4; 6-8; 27; 32; 33; 44; 46], but the impact of pain on higher-level cognitive processes is less-investigated (although see [1,23,28]). It is unclear whether we should expect higher-level tasks to also be affected by pain. Tasks which draw on multiple attention processes and executive functions which are themselves disrupted by pain may see an additive disruptive effect. Alternatively, performance on these tasks may be protected. It is possible that some of the attention processes involved are able to compensate for the disruption to others. It is also possible that the tasks are more engaging for participants which motivates them to focus their attention on the task rather than the pain [45].

Here, we investigated the effect of pain on logical reasoning. Reasoning encompasses many domains, but can broadly be thought of as the process of drawing inferences from available information. Reasoning is often understood in terms of dual processes. Type 1 processing is autonomous whereas Type 2 processing requires working memory [10; 15]. While Type 1 processing is highly efficient and effective, many tasks in modern life require Type 2 processing, in particular, novel, abstract and complex tasks such as choosing the best insurance plan or deciding whether one is eligible for benefit schemes [37]. Such tasks might require calculating and comparing costs and probabilities, or interpreting conjunctions of conditional statements, and Type 2 processing is essential for performing these tasks optimally. Any factors that systematically reduce the engagement of Type 2 processes during reasoning may, therefore, impact on the quality of life of the individual. Factors known to reduce Type 2 thinking on reasoning tasks include time pressure [12; 13; 22] and working memory load [20; 21]. However, Type 2 processing can also be increased by instruction variations [14] and prompts that indicate task difficulty [5].

This is the first investigation of the effect of pain on logical reasoning that we are aware of, and we began by selecting three tasks that measure different types of reasoning: the Cognitive Reflection Test (CRT[19]), the Belief Bias Syllogisms Task [37], and the Conditional Inference Task [9]. We took the approach of investigating three tasks, rather than focusing on one, because reasoning is an inherently broad concept. The CRT questions prompt intuitive but incorrect responses (resulting from Type 1 processing), the error of which can be spotted and corrected on reflection (with Type 2 processes). Responses therefore indicate a participant's tendency to spontaneously engage Type 2 thinking. The syllogisms task requires participants to set aside their prior beliefs (which come from Type 1 processes) and decide whether conclusions follow logically (requiring Type 2 processing) from syllogisms that are framed in believable, unbelievable or neutral context. The Conditional Inference task requires participants to decide whether conclusions follow logically from abstract conditional statements, making it a 'purer' measure of logical reasoning ability, free from context. We predicted that pain would reduce logical responding on all three tasks.

## 2. Experiment 1

### 2.1.Method

#### 2.1.1. Participants

Participants were 60 students and staff (29 males, 31 females) from the University of Bath and Bath Spa University, aged 18-58 years ($M$ = 26.17, $SD$ = 8.73). All reported being free from pain at the time of the study.

#### 2.1.2. Design

Participants completed three reasoning tasks and we investigated the effect of pain on performance on each task. The Cognitive Reflection Test was administered first, with pain condition as a counterbalanced between-participants factor. Participants then completed two forms of the Belief Bias Syllogisms and Conditional Inference tasks, one form of each while experiencing heat pain and one form of each while pain-free (i.e. within-groups). The order of tasks and pain conditions was counterbalanced, with participants completing one of the following four orders: 1) syllogisms pain, syllogisms no-pain, conditionals pain, conditionals no-pain, 2) syllogisms no-pain, syllogisms pain, conditionals no-pain, conditionals pain, 3) conditionals pain, conditionals no-pain, syllogisms pain, syllogisms no-pain, 4) conditionals no-pain, conditionals pain, syllogisms no-pain, syllogisms pain.

### 2.1.3. Measures

#### 2.1.3.1. Cognitive Reflection Test

Participants completed the original three-item Cognitive Reflection Test (CRT[19]) plus the four new items developed by Toplak, West and Stanovich [41] on paper, either with or without pain. For each item there is an intuitive but incorrect response. If a participant inhibits this intuitive response, they can fairly easily calculate the correct response. For example, in the well-known bat and ball problem from the original three items, "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?", the intuitive answer is 10 cents while the correct answer is 5 cents.

The CRT is a popular measure in dual-processes research because it is thought to indicate an individual's likelihood to override Type 1 processing (which results in the intuitive response) with Type 2 processing (which in the case of the CRT usually leads to the correct response). Toplak, West and Stanovich [40] found the CRT to be a better predictor of normative responding to reasoning tasks than cognitive ability, executive function, or the

Actively Openminded Thinking scale, supporting its utility. The CRT was administered to each participant once, with or without pain, with conditions counterbalanced.

2.1.3.2. Belief Bias Syllogisms Task

The Belief Bias Syllogisms task [37] was used as a measure of the ability to reason independently of prior beliefs. The original 24 items were split into two equivalent 12-item forms [3] to allow us to give participants opposite forms in each pain condition. The order of forms and conditions was counterbalanced.

Each form consisted of 12 thematic syllogisms, four congruent (believable-valid, unbelievable-invalid), four incongruent (believable-invalid, unbelievable-valid) and four neutral (nonsense-valid, nonsense-invalid). An invalid syllogism with a believable conclusion is "All living things need water; Roses need water; Therefore, roses are living things". Participants decided whether each syllogism was logically valid or not after being instructed to ignore their prior beliefs. We measured the number of items endorsed for each item type (believable valid, believable invalid etc.) in each condition.

2.1.3.3. Conditional Inference Task

Participants completed 16 items from Evans, Clibbens and Rood's version of the Conditional Inference Task[9]. The task consists of abstract inferences of four forms: Modus Ponens (MP: if p then q; p; q), Denial of the Antecedent (DA: if p then q; not p; not q), Affirmation of the Consequent (AC: if p then q; q; p) and Modus Tollens (MT: if p then q; not q; not p). MP and MT inferences are considered valid and DA and AC inferences considered invalid under the normative model of the conditional inference task. An example Modus Tollens item is: "If the letter is K then the number is 5; The number is not 5; Therefore, the letter is not K". Participants completed the 16 items that had explicit rather

than implicit negations in the premises (i.e. “The number is not 5” rather than “The number is 7” in the example above).

The lexical content of the rules was generated randomly and varied between two forms of the task. Each participant performed one form while in pain and the other while pain free, with pain condition and form counterbalanced. The task was presented on E-Prime Professional 2.0 with the order of items randomized for each participant. Participants responded to each item by pressing the leftmost button on a serial response box to indicate that the 'conclusion follows', or the rightmost button on the response box to indicate that the 'conclusion does not follow'. The task was preceded by the instructions used by Evans et al (1995).

2.1.3.4. VAS scales

At the end of the study, participants completed 10 cm Visual Analogue Scales on paper for the following questions: "How painful was the heat?" (anchored by "No pain at all" and "Unbearable pain"), "How much pain did you feel when the heat pain was at its most intense?" (anchored by "No pain at all" and "Unbearable pain"), and "How intrusive/distracting did the pain seem to you?" (anchored by "Not at all distracting" and "Very distracting").

2.1.3.5. Pain induction

A Medoc Pathway Advanced Thermal Stimulator (ATS) was used for the purpose of heat pain induction. This equipment is designed for use in clinical and research settings and has built-in safety restrictions. The pain induction protocol began by determining the participant's pain threshold, and the subsequent pain induction during the reasoning tasks was tailored to this threshold.

The thermode, measuring 30 × 30mm, was attached to the participant's non-dominant arm over the extensor digitorum muscle, approximately 5cm above the wrist joint. The baseline temperature of the thermode was set at 32°C and participants were asked to increase the temperature using the manual trigger until it first felt painful (i.e. reached the pain threshold). If the heat remained painful but tolerable for 15 seconds, this temperature was recorded as the threshold. Otherwise, the temperature was adjusted until this criterion was satisfied. Once the threshold was determined the heat stimulus was returned to the baseline temperature. This procedure was repeated three times, and the mean threshold temperature was taken as the participants' heat pain threshold.

The pain induction procedure during the tasks was a pulses programme in which the thermode temperature fluctuated between 1°C below and above the participant's pain threshold in cycles of 10. Each cycle of 10 pulses lasted for 5 seconds. Between each cycle, there was a 3 second return to baseline, before the next cycle of 10 fluctuations around the threshold. After a participant's pain threshold had been determined, but before they started the tasks, they experienced the pulses programme tailored to their threshold for 30 seconds and were asked to verbally report whether or not the heat was painful but tolerable. Eighteen participants reported that the pulses programme was not painful, and the temperatures were then increased one degree at a time and the 30-second programme repeated until the participant reported that they experienced pain. Three participants reported that the pulses programme was too painful, and the temperatures were then decreased one degree at a time and the 30-second programme repeated until the participant reported that they experienced tolerable pain. The remaining 39 participants reported that the pulses programme was painful but tolerable on the initial demonstration.

### 2.1.4. Procedure

Ethical approval for the study was granted from the University of Bath Department of Psychology and Department for Health ethics boards. After giving informed consent, the participants' heat pain threshold was determined as described above. Next, participants completed the CRT task with or without pain, then the syllogisms and conditional inference tasks twice, once while pain free and once while experiencing the painful heat stimulus.

After completing the reasoning tasks participants completed the VAS scales on paper and some demographics questions on the computer using E-Prime Professional 2.0. When all tasks had been completed the participants were paid and debriefed.

## 2.2. Results

The key frequentist analyses below are followed up by Bayesian analyses.

For ANOVAs with two or fewer within-subjects factors we report Generalized Eta Squared ($\eta_G^2$) as a measure of effect size, as recommended by Lakens [29], in addition to Partial Eta Squared ($\eta_p^2$). For analyses with more than two within-subjects factors, we report only $\eta_p^2$ [29].

### 2.2.1. Data availability

The data and SPSS analysis scripts for these experiments are available at https://figshare.com/s/a55106b572b7c4a1329d.

### 2.2.2. Data cleaning

One participant's syllogisms data and another participant's CRT data were removed due to an error in administration.

### 2.2.3. Pain threshold and VAS ratings

The mean pain threshold was 43.68°C (*SD* = 3.35), and was significantly higher in males (*M* = 44.68°C, *SD* = 2.94) than in females (*M* = 42.67°C, *SD* = 3.46), *t*(58) = 2.43, *p* = .018, *d* = .62. The mean VAS response to the question "How much pain did you feel during the pain condition of the task?" was 58.12 out of 100 (*SD* = 16.52), ranging from 17 to 88, which was significantly different from zero, *t*(59) = 27.25, *p* < .001, *d* = 3.52. The mean response to the question "How much pain did you feel when the heat pain was at its most intense?" was 68.67 (*SD* = 14.55), which was also significantly different from zero, *t*(59) = 36.56, *p* < .001, *d* = 4.72. Finally, the mean response to the question "How intrusive/distracting did the pain seem to you?" was 57.72 (*SD* = 23.49), which again was significantly different from zero, *t*(59) = 19.03, *p* < .001, *d* = 2.46. None of the VAS ratings differed by sex (all *ps* > .13). These results suggest that participants perceived the heat stimulus to be painful and that they perceived this pain to be distracting.

### 2.2.4. The effect of pain on task performance

#### *2.2.4.1. CRT*

Given that the CRT7 includes four new items which are relatively untested, we examined scores on the original three items and well as scores on the full seven items. The number of correct answers for each measure was compared across the pain and no pain conditions with Mann-Whitney U tests. For CRT3 scores, there was no difference between the pain (*M* = .97, *SD* = 1.12) and no pain (*M* = 1.27, *SD* = 1.20) conditions, *U*(59) = 372.5, *z* = 1.00, *p* = .318, *r* = .130. For CRT7 scores, there was again no difference between the pain (*M* = 2.62, *SD* = 2.09) and no pain (*M* = 3.30, *SD* = 2.20) conditions, *U*(59) = 353, *z* = 1.26, *p* = .208, *r* = .164.

We also conducted Bayesian Mann-Whitney U tests on the effects of pain on CRT3 and CRT7 scores with a default Cauchy prior of 0.707. For CRT3 scores there was anecdotal

evidence in favour of the null, BF_01 = 2.65, and for CRT7 scores there was anecdotal evidence in favour of the null, BF_01 = 1.83.

Within the pain group, there was a significant negative correlation between CRT3 scores and self-reported average pain intensity, $r_s(29) = -.508$, $p = .005$, and a significant negative correlation between CRT7 scores and self-reported average pain intensity, $r_s(29) = -.462$, $p = .012$; the more intense the pain, the fewer correct answers participants gave. Bayesian correlations between pain intensity scores and CRT3 scores, and between pain intensity scores and CRT7 scores, indicated strong evidence in favour of the alternative hypotheses, BF_10 = 56.61 and BF_10 = 18.16, respectively.

*2.2.4.2. Belief Bias Syllogisms*

Endorsement rates for the syllogisms were entered into a 2 (pain condition: pain, no pain) × 3 (believability: believable, neutral, unbelievable) × 2 (validity: valid, invalid) × 2 (sex: female, male) ANOVA. Importantly, this replicated the usual findings associated with the task. There was a significant main effect of believability, $F(2,114) = 20.10$, $p < .001$, $\eta_p^2 = .261$. As expected, participants endorsed believable conclusions ($M = 76.3\%$, $SD = 22.5$) more often than neutral conclusions ($M = 67.6\%$, $SD = 23.2$), $p = .001$, and neutral conclusions more often than unbelievable conclusions ($M = 57.4\%$, $SD = 27.9$), $p = .001$. There was also a significant main effect of validity, $F(1,57) = 77.18$, $p < .001$, $\eta_p^2 = .575$. Participants endorsed valid conclusions ($M = 86.6\%$, $SD = 14.7$) more often than invalid conclusions ($M = 47.6\%$, $SD = 34.9$), again as expected.

There was a significant interaction between believability and validity, $F(2,114) = 11.26$, $p < .001$, $\eta_p^2 = .165$ (Figure 1). Further investigation showed that there was a main effect of believability on endorsement rates both within the valid items, $F(2,116) = 11.49$, $p < .001$, $\eta_p^2 = .165$, and within the invalid items, $F(2,116) = 22.95$, $p < .001$ $\eta_p^2 = .284$. There

were significant differences between believable, neutral and unbelievable items within both the valid and invalid items, all *ps* < .041. Notably, believable items were highly endorsed, more often than would be expect by chance, whether they were valid (*M* = 91.1%, *SD* = 14.7), *p* < .001 or invalid (*M* = 61.4%, *SD* = 36.4), *p* = .018. Unbelievable items, however, were endorsed more often than would be expect by chance when they were valid, (*M* = 77.5%, *SD* = 29.4), *p* < .001, but less often than would be expected by chance when they were invalid, (*M* = 37.3%, *SD* = 38.0), *p* = .012. This replicates findings from previous investigations of the task, which are interpreted as 'selective scrutiny' of unbelievable inferences[10]. Believable inferences tend to be highly endorsed whether they are valid or invalid, but unbelievable inferences are subjected to extra scrutiny with Type 2 processing, which leads to them being endorsed when they are valid and rejected when they are invalid.

There were no main effects or interactions involving pain or sex (all *p*s > .052, $\eta_p^2$ < .065).

We also conducted a Bayesian ANOVA on Syllogisms endorsement rates with Believability, Validity, Pain Condition and Sex as factors. We used Rouder et al.'s [36] default prior width of h = 0.5, which captures the range of effect sizes typically found in the behavioural sciences. This revealed that the best-fitting model included Belief, Validity and a Belief × Validity interaction, P(model|data) = .52. There was moderate evidence against the model that additionally included pain, BF_01 = 5.77, and very strong evidence against the model that included pain, its interactions with Belief and Validity, and the three-way interaction, BF_01 = 10024.

Given that the level of pain experienced varied widely between participants (from 17 to 88 on a VAS scale of 0 to 100) we ran some exploratory analyses to investigate the effect of pain intensity on performance in the pain condition. For responses in the pain condition only, we calculated a Total Correct score and a Belief Bias Index, which reflects participants'

tendency to reason based on believability over logical validity ([number of believable valid items endorsed + number of unbelievable invalid items endorsed] – [number of believable invalid items endorsed + number of unbelievable valid items endorsed]) [37]. Total Correct scores were not correlated with Pain Intensity, $r(59) = -.176$, $p = .183$, but Belief Bias Index scores were significantly positively correlated with Pain Intensity, $r_s(59) = .345$, $p = .007$. This suggests that the higher a participant's pain intensity, the more swayed they were by the believability of the syllogisms.

A Bayesian Pearson's correlation between Pain Intensity and Total Correct syllogisms score found anecdotal evidence in favour of the null hypothesis, BF_01 = 2.589. A Bayesian Kendall's tau-b correlation between Pain Intensity and Belief Bias Index found moderate evidence in favour of the alternative hypothesis, BF_10 = 15.19.

*2.2.4.3. Conditional Inference Task*

Conditional inference endorsement rates were entered into a 2 (condition: pain, no pain) x 4 (inference: MP, DA, AC, MT) x 2 (sex: male, female) ANOVA. There was a significant main effect of Inference, $F(3,174) = 54.61$, $p < .001$, $\eta_p^2 = .485$, $\eta_G^2 = .275$, with MP inferences being endorsed more often ($M = 98.1\%$, $SD = 5.1$) than AC inferences ($M = 73.5\%$, $SD = 33.8$), $p < .001$, which in turn were endorsed more often than MT inferences ($M = 65.0\%$, $SD = 26.9$), $p = .029$, which were endorsed more often than DA inferences ($M = 52.7\%$, $SD = 25.0$), $p = .002$. There were no main effects of pain or sex nor any interactions (all $p$s > .246, $\eta_p^2 < .024$, $\eta_G^2 < .006$).

We conducted a Bayesian ANOVA with Inference, Pain Condition and Sex as factors. We used Rouder et al.'s [36] default prior width of h = 0.5, which captures the range of effect sizes typically found in the behavioural sciences. This revealed that the best-fitting model included only Inference type, P(model|data)=.67. There was moderate evidence against the

model that additionally included pain, BF_01=8.06, and very strong evidence against the model that included Pain, Sex and all associated interactions, BF_01=1,525,000.

Next, as with the Syllogisms scores, we ran exploratory analyses investigating the relationships between pain intensity and conditional inference endorsement rates in the pain condition only. Pain Intensity was not correlated with the number of items endorsed for any inference type: MP $r_s(60) = -.166$, $p = .205$, DA $r_s(60) = .231$, $p = .076$, AC $r_s(60) = .234$, $p = .071$, MT $r_s(60) = .185$, $p = .158$.

### 2.3. Experiment 1 Discussion

Experiment 1 did not support our primary hypothesis that participants' performance on the reasoning tasks would be poorer when they were in pain compared to when they were pain-free. However, some exploratory analyses suggested that higher intensity pain may affect reasoning more than low intensity pain. We observed a negative correlation between pain intensity and the number of CRT items answered correctly. On the syllogisms task, the higher a participant's self-reported pain intensity was, the more their responses relied on the believability of the syllogisms over their logical validity.

These exploratory analyses do not provide evidence for an effect of pain on reasoning, but they suggest that our failure to support our hypotheses may have been due to our pain manipulation being inadequate. In Experiment 2, participants completed the CRT and Syllogisms task again, but we improved our pain manipulation by increasing the thermode temperature to ensure that all participants experienced moderate pain. We predicted that this would lead to a main effect of pain on CRT scores, and an interaction between pain, validity, and believability on the Syllogisms task. We did not include the Conditional Inference Task because we wanted to make the testing sessions shorter due to the higher intensity pain induction, and the theoretical reasons to expect performance on this task to be affected by

pain were not as strong as for the CRT and Syllogisms tasks, on which responding patterns more clearly differ depending on the extent of Type 1 vs Type 2 thinking participants used. On the Conditional Inference Task, it is less clear exactly how response patterns would change if pain affected participants' reasoning – there is no clear "default" choice between responding "yes" or "no" when relying on Type 1 thinking.

## 3. Experiment 2

### 3.1. Method

#### 3.1.1. Participants

Participants were 46 students and staff (20 males, 26 females) from the University of Bath and Bath Spa University, aged 18-34 years ($M$ = 22.98, $SD$ = 4.59). All reported being free from pain at the time of the study.

#### 3.1.2. Design

Participants completed the CRT and Belief Bias Syllogisms tasks and we investigated the effect of pain on performance on each task. The Cognitive Reflection Test was administered first, with pain condition as a counterbalanced between-participants factor. Participants then completed the Belief Bias Syllogisms task twice, once while experiencing heat pain and once while pain free. The order of pain conditions was counterbalanced.

#### 3.1.3. Pain induction

A Medoc Pathway Advanced Thermal Stimulator (ATS) was again used for heat pain induction. The same pulses programme was used as in Study 1, where the temperature increased and fluctuated by 2°C for 5 seconds, with 3-second returns to the baseline temperature of 32°C, before pulsing at the higher temperatures again. However, in this study,

the pulsing temperatures were set to at least 44°C and 46°C in the pain condition, with the same baseline of 32°C. At the start of the experiment, participants rated how painful they found this on a Numerical Rating Scale (NRS) where 0 was labelled 'no pain at all' and 10 was labelled 'unbearable pain'. If the participant rated it as lower than 7/10 on the NRS, the temperatures were increased by 1°C and the NRS repeated. This continued until a temperature was found that participants rated as 7/10 or higher, using increments or decrements of half a degree where necessary to find a bearable temperature rated at least 7/10. In the warm condition, the pulsing temperatures were set at 39°C and 41°C and the baseline was again 32°C. This gave participants a similar sensation of fluctuating temperatures to the pain condition, but without pain.

### 3.1.4. Measures

Participants again completed the seven-item CRT[19; 41], the two 12-item forms of the Belief Bias Syllogisms task[37], and the same VAS scales assessing average pain, maximum pain, and pain intrusiveness.

### 3.1.5. Procedure

Ethical approval for the study was granted from the University of Bath Department of Psychology and Department for Health ethics boards. After giving informed consent, the temperature of the thermode was chosen for each participant to meet the criteria of being at least 44°C and 7/10 on a NRS. Next, participants completed the CRT task with or without pain, then the syllogisms task twice, once with warm but non-painful heat and once with painful heat. The order of pain conditions was counterbalanced.

After completing the reasoning tasks participants completed the VAS scales on paper and some demographics questions on the computer using E-Prime Professional 2.0. When all tasks had been completed the participants were paid and debriefed.

## 3.2. Results

### 3.2.1. Data availability

The data and SPSS analysis scripts for these experiments are available at https://figshare.com/s/a55106b572b7c4a1329d.

### 3.2.2. VAS ratings

The mean VAS response to the question "How much pain did you feel during the pain condition of the task?" was 62.75 out of 100 (*SD* = 11.72), ranging from 28 to 87, which was significantly different from zero, $t(45) = 36.33$, $p < .001$, $d = 5.35$, but did not differ between males (M = 61.83, SD = 8.28) and females (M = 63.46, SD = 13.92), $t(44) = .47$, $p = .644$, $d = .142$. The mean response to the question "How much pain did you feel when the heat pain was at its most intense?" was 76.33 (*SD* = 10.94), which was also significantly different from zero, $t(45) = 47.31$, $p < .001$, $d = 6.98$, and did not differ between males (M = 75.15, SD = 9.67) and females (M = 77.23, SD = 11.94), $t(44) = .635$, $p = .529$, $d = .191$. Finally, the mean response to the question "How intrusive/distracting did the pain seem to you?" was 56.07 (*SD* = 19.68), which was significantly different from zero, $t(45) = 19.32$, $p < .001$, $d = 2.85$, and significantly higher in females (*M* = 61.12, *SD* = 17.94) than in males (*M* = 49.50, *SD* = 20.34), $t(44) = 2.05$, $p = .046$, $d = .606$. Overall, these results suggest that participants perceived the heat stimulus to be painful and that they (particularly females) perceived this pain to be distracting.

3.2.3. The effect of pain on task performance

*3.2.3.1. CRT*

We again examined scores on the original three CRT items and well as scores on the full seven items. The number of correct answers for each measure was compared across the pain and no pain conditions with Mann-Whitney U tests. For CRT3 scores, there was no difference between the pain ($M$ = .83, $SD$ = .1.05) and no pain ($M$ = 1.10, $SD$ = .97) conditions, $U(46) = 217$, $z$ = -1.10, $p$ = .273, $r$ = .162. For CRT7 scores, there was again no difference between the pain ($M$ = 2.46, $SD$ = 1.89) and no pain ($M$ = 3.23, $SD$ = 1.80) conditions, $U(46) = 192.5$, $z$ = -1.601, $p$ = .109, $r$ = .236.

Bayesian Mann-Whitney U tests on the effects of pain on CRT3 and CRT7 scores, with a default Cauchy prior of 0.707, found anecdotal evidence in favour of the null hypothesis for both CRT3, BF_01 = 2.30, and CRT7 scores, BF_01 = 1.71.

Within the participants who completed the CRT in pain, there was no correlation between CRT3 scores and average pain intensity, $r_s$ (24) = .375, $p$ = .071, nor between CRT7 scores and average pain intensity, $r_s$ (24) = .251, $p$ = .237.

Bayesian correlations between pain intensity scores and CRT3 scores, and between pain intensity scores and CRT7 scores, indicated weak evidence in favour of the alternative hypotheses for CRT3 scores, BF_10 = 3.11, and anecdotal evidence in favour of the null hypothesis for CRT7 scores, BF_01 = 2.69.

*3.2.3.2. Belief Bias Syllogisms*

Endorsement rates for the syllogisms were entered into a 2 (pain condition: pain, no pain) × 3 (believability: believable, neutral, unbelievable) × 2 (validity: valid, invalid) × 2 (sex: male, female) ANOVA. There was a significant main effect of believability, $F(2,88)$ = 25.65, $p < .001$, $\eta_p^2$ = .368, where participants endorsed believable conclusions ($M$ = 71.5%,

*SD* = 21.05) more often than neutral conclusions (*M* = 61.1%, *SD* = 19.0), $p < .001$, and neutral conclusions more often than unbelievable conclusions (*M* = 49.6%, *SD* = 23.1), $p < .001$. There was a significant main effect of validity, $F(1,44) = 112.67$, $p < .001$, $\eta_p^2 = .719$, where participants endorsed valid conclusions (*M* = 85.8%, *SD* = 13.6) more often than invalid conclusions (*M* = 35.6%, *SD* = 30.5), again as expected. There was also a significant main effect of Sex, $F(1,44) = 4.30$, $p = .044$, $\eta_p^2 = .089$, where women endorsed more conclusions (*M* = 66.0%, *SD* = 17.3) than men (*M* = 55.4%, *SD* = 17.0).

There was a significant interaction between believability and validity, $F(2,88) = 8.61$, $p < .001$, $\eta_p^2 = .164$. Further investigation showed that there was a main effect of believability on endorsement rates both within the valid items, $F(2,90) = 11.33$, $p < .001$, $\eta_p^2 = .201$, and within the invalid items, $F(2,90) = 27.58$, $p < .001$ $\eta_p^2 = .380$. For the valid items, there was no significant difference in endorsement rates for believable (*M* = 91.3%, *SD* = 14.9) and neutral items (*M* = 89.7%, *SD* = 16.3), $p = .519$, but both differed significantly from unbelievable items (*M* = 76.6%, *SD* = 23.1), both *ps* = .001. For the invalid items, believable items were endorsed more often (*M* = 53.3%, *SD* = 36.6) than neutral items (*M* = 34.2%, *SD* = 37.3), $p < .001$, which in turn were endorsed more often than unbelievable items (*M* = 23.4%, *SD* = 32.6), $p = .008$. Notably, valid items were highly endorsed whether they were believable, neutral or unbelievable (with endorsement rates significantly higher than chance level for all three, all *p*s < .001), whereas invalid items were endorsed less often than not only when they were neutral ($p = .006$) or unbelievable ($p < .001$). Contrary to logic, believable but invalid items were not rejected more often than would be expected by chance, $p = .550$. Again, this replicates findings from previous investigations of the task, which are attributed to selective scrutiny of unbelievable syllogisms.

There was a significant interaction between Validity and Sex, $F(1,44) = 4.30$, $p = .044$, $\eta_p^2 = .089$, where females endorsed significantly more invalid items (*M* = 45.8%, *SD* =

28.4) than males ($M$ = 25.4%, $SD$ = 32.8), $t(44) = 2.26$, $p = .029$, while there was no difference in the number of valid items endorsed between males ($M$ = 85.4%, $SD$ = 15.7) and females ($M$ = 86.2%, $SD$ = 10.8), $t(44) = .205$, $p = .839$.

There was no main effect of pain on endorsement rates, $F(1,44) = 1.82$, $p = .184$, $\eta_p^2 = .040$, however, there was a significant interaction between Pain, Validity and Believability, $F(2,88) = 3.44$, $p = .036$, $\eta_p^2 = .073$. To locate the source of this three-way interaction, we ran a Pain × Validity ANOVA separately for the believable, neutral and unbelievable items. For believable items, there was a significant interaction between Pain and Validity $F(1,45) = 5.37$, $p = .025$, $\eta_p^2 = .107$ (see Figure 2). Although not significant at the simple main effect level, endorsement rates for invalid items were lower when participants were in pain ($M$ = 47.8%, $SD$ = 39.3) compared to not in pain ($M$ = 58.7%, $SD$ = 44.1), $t(45) = 1.88$, $p = .067$, whereas endorsement rates for valid items were higher when participants were in pain ($M$ = 93.5%, $SD$ = 17.0) compared to when they were not in pain ($M$ = 89.1%, $SD$ = 21.0), $t(45) = 1.27$, $p = .209$ (i.e. a cross-over interaction). There was no interaction between Pain and Validity for either neutral, $F(1,45) = 2.46$, $p = .124$, $\eta_p^2 = .052$, or unbelievable items, $F(1,45) < .001$, $p = 1.000$, $\eta_p^2 < .001$.

All other effects were non-significant (all $p$s > .318, $\eta_p^2 < .023$).

We also conducted a Bayesian ANOVA on syllogisms endorsement rates with Believability, Validity, Pain Condition and Sex as factors. Again, we used Rouder et al.'s [36] default prior width of h = 0.5. This revealed that the best-fitting model included Believability, Validity, Sex, Believability × Validity and Sex × Validity, P(model|data) = .49. There was weak evidence against the model that additionally included pain, BF_01=3.80, and very strong evidence against the model that additionally included Pain and all interactions with Pain, BF_01= 3133.

### 3.3. Experiment 2 Discussion

In Experiment 2, where the pain condition was made more painful, pain seemed to improve performance on believable items by making participants more likely to endorse valid items and less likely to endorse invalid items, compared to the pain-free condition. However, this effect differs from Experiment 1 and was not predicted. The key methodological difference between Experiments 1 and 2 was that we ensured that the participants experienced a pain intensity of at least 7/10 in Experiment 2, which we did not do in Experiment 1. We had expected this higher-intensity pain to disrupt participants' performance, but finding the opposite led us to hypothesise that pain above a certain intensity may actually cue participants to increase their effort in order to protect their task performance.

In Experiment 3, we directly manipulated pain intensity within-subjects to allow us to test this hypothesis. Participants completed CRT and syllogisms items in no pain, low-intensity pain and high-intensity pain conditions. We predicted that performance would decline in the low-intensity pain condition relative to the no pain condition, and improve in the high-intensity pain condition compared to the low-intensity pain condition.

## 4. Experiment 3

### 4.1. Method

#### 4.1.1. Participants

Participants were 38 students and staff (27 females, 9 males, 2 missing data due to technical error) from the University of Bath and Bath Spa University, aged 19-53 years ($M$ = 27.54, $SD$ = 6.23). All reported being free from pain at the time of the study. Due to the small number of male participants, we did not explore sex differences in this study.

### 4.1.2. Design

Participants completed versions of the CRT and Belief Bias Syllogisms tasks that were altered to allow for three within-groups pain conditions (described below). We also included two additional measures: a short version of Raven's Advanced Progressive Matrices, and confidence ratings after each syllogism. If pain had any effects on the Syllogisms or CRT tasks, we wanted to be able to see whether it was mediated by changes to confidence (self-rated) or cognitive capacity (measured by Raven's Matrices). For example, if high intensity pain did improve reasoning performance by acting as a cue to task difficulty, as predicted, participants may feel less confident in their syllogism responses in the high intensity pain condition compared to the low intensity condition, and should perform better on the Raven's Matrices task as well. Pain intensity was directly manipulated in order to experimentally investigate the effect of pain intensity on task performance. Each task was performed in three within-groups conditions: no pain, low intensity pain, and high intensity pain.

### 4.1.3. Pain induction

A Medoc Pathway Advanced Thermal Stimulator (ATS) was again used for heat pain induction, but in this experiment the thermode was placed on the participant's left ankle instead of their wrist. This was because participants were invited to take part in another pain-related experiment after this one, and those who did experienced pain to their hand and wrist via a cold pressor. We aimed to minimise the exposure of any one body area to pain induction procedures over the course of the two experiments.

The procedure began by finding each participants' pain threshold, using the procedure from Study 1. Participants increased the thermode temperature until it first became painful and remained painful but tolerable for 15 seconds. An average of three trials was taken as the

participant's threshold. This temperature was then applied to the pulses programme, and the participant experienced this for 30 seconds then rated the pain intensity on a 0-10 NRS, where 0 was labelled 'no pain', 2 was labelled 'mild pain', 4-5 was labelled 'moderate pain', 7-8 was labelled 'high pain', and 10 was labelled 'maximum tolerable pain'.

In order to identify temperatures for the low and high pain intensity conditions, the temperature was then adjusted using a staircase method, first going 2 degrees up, then 1 degree down, with the participant experiencing the pulses programme for 30 seconds at each temperature and rating the pain intensity on the 0 – 10 NRS. This continued until an NRS rating of 8 or 9 was reached (whichever was reached first). The temperatures rated as 2 out of 10 (or 3 if the staircase method did not produce a rating of 2) and 7 out of 10 (8 if the staircase method did not produce a rating of 7) were selected for the low and high pain intensity conditions, respectively. In the no pain condition, the baseline temperature of 32 degrees was applied.

#### 4.1.4. Measures

##### *4.1.4.1. Cognitive Reflection Test*

Contrary to Experiments 1 and 2, the CRT was administered in the three pain conditions within-groups, by splitting it into three 2-item blocks (items counterbalanced). This allowed us to experimentally investigate the effect of pain intensity on performance. Six items were selected using the performance data given by Toplak, West & Stanovich [39]: items 1-6 were included and item 7 (the stocks question) was excluded because it showed the lowest correlation with the original three items.

##### *4.1.4.2. Belief Bias Syllogisms*

We used a different version of the syllogisms task for Experiment 3, based on recent research by Trippas and colleagues [42; 43]. This version allowed for more items per-condition and had previously been used alongside confidence ratings. After each question, participants were asked to rate their confidence in their answer from 1 (not confident) to 3 (very confident). Participants saw 16 different syllogisms in each of the no pain, low pain and high pain conditions, with the same item structures but different content in each condition. Half of the 16 syllogisms in each condition were valid and half invalid. Half had believable conclusions and half had unbelievable conclusions (there were no belief-neutral items). Premise believability was controlled for using pseudo-word middle terms, so that the belief or disbelief stemmed only from the conclusion (e.g. No birds are pinds; Some pinds are parrots; Some parrots are not birds).

#### *4.1.4.3. Raven's Matrices*

An 18-item subset of items from Raven's Advanced Progressive Matrices [35] was used as a measure of cognitive capacity [37; 39]. The 18 items were split into three lists of six items (list one: items 1, 4, 7, 10, 13, 16; list two: items 2, 5, 8, 11, 14, 17; and list three; items 3, 6, 9, 12, 15, 18). The lists were counterbalanced by pain condition. Six items cannot be considered a comprehensive measure of cognitive capacity, rather, this was intended as a proxy measure given the time constraints of pain induction procedures.

#### *4.1.4.4. VAS ratings*

After the low and high pain conditions, participants completed the same three VAS ratings used in Experiments 1 and 2.

### 4.1.5. Procedure

Ethical approval for the study was granted from the University of Bath Department of Psychology and Department for Health ethics boards. After giving informed consent, the temperature of the thermode was chosen for each participant for each condition as described above. Next, participants completed practice blocks of the CRT, Raven's and Syllogisms tasks without pain. Participants then completed the main blocks of the three tasks under the no pain, low pain and high pain conditions. The order of these conditions was counterbalanced by E-Prime, and the order of the three tasks within each pain condition was also counterbalanced by E-Prime. Participants completed the VAS scales on paper at the end of the low and high intensity pain segments. When all tasks had been completed the participants were paid and debriefed. The majority of participants then remained in the laboratory to take part in a separate study.

## 4.2. Results

### 4.2.1. Data availability

The data and SPSS analysis scripts for these experiments are available at https://figshare.com/s/a55106b572b7c4a1329d.

### 4.2.2. VAS ratings

The mean VAS response to the question "How painful was the heat on average?" was 26.10 (*SD* = 17.79) in the low pain condition and 67.45 (*SD* = 18.15) in the high pain condition. These significantly differed from each other, $p < .001$, $d = 2.30$, and from zero, both *p*s < .001, both *d*s > 1.47. The mean response to the question "How much pain did you feel when the heat pain was at its most intense?" was 30.26 (*SD* = 20.29) in the low pain condition and 76.77 (*SD* = 17.81) in the high pain condition. These significantly differed from each other, $p < .001$, $d = 2.44$ and from zero, both *p*s < .001, both *d*s > 1.49. Finally, the

mean response to the question "How intrusive/distracting did the pain seem to you?" was 24.87 (*SD* = 19.84) in the low pain condition and 67.52 (*SD* = 22.29) in the high pain condition. These significantly differed from each other, $p < .001$, $d = 2.02$, and from zero, both *p*s < .001, both *d*s > 1.25. These results suggest that the pain manipulations were successful and that the pain was subjectively distracting in both conditions.

4.2.3. The effect of pain on task performance

*4.2.3.1. CRT*

A Friedman's Test found no effect of pain condition on the number of CRT items answered correctly, $\chi^2$ (2) = .286, $p = .867$. Effect sizes for Friedman's test can be calculated from Wilcoxon's tests on the paired comparisons [17], which in this case were all small (*r*s < .092).

A Bayesian ANOVA on CRT scores with Pain Condition as a factor and Rouder et al.'s [36] default prior width of h = 0.5 revealed that the best-fitting model was the null model, P(model|data) = .91. There was moderate evidence against a model that additionally included Pain, BF_01 = 9.99.

*4.2.3.2. Raven's Matrices*

An ANOVA found no effect of pain condition on the number of Raven's Matrices answered correctly, $F(2,74) = 1.33$, $p = .270$, $\eta_p^2 = .035$, $\eta G^2 = .010$.

A Bayesian ANOVA examining the effect of pain condition on the number of Raven's Matrices answered correctly using Rouder et al.'s [36] default prior width of h = 0.5 found that the best-fitting model was the null model, P(model|data) = .80. There was moderate evidence against a model that additionally included Pain, BF_01 = 4.09.

*4.2.3.3. Belief Bias Syllogisms (endorsement rates)*

A 3 (Pain: none, low-intensity, high-intensity) × 2 (Validity: Valid, Invalid) × 2 (Believability: Believable, Unbelievable) ANOVA on syllogism endorsement rates showed a significant main effect of Validity, $F(1,37) = 59.10$, $p < .001$, $\eta_p^2 = .615$, with higher endorsement of valid inferences ($M = 75.7\%$, $SD = 15.4$) than invalid inferences ($M = 53.5\%$, $SD = 17.9$). There was a significant main effect of Believability, $F(1,37) = 20.80$, $p < .001$, $\eta_p^2 = .360$, with higher endorsement of believable items ($M = 74.3\%$, $SD = 17.9$) than unbelievable items ($M = 55.0\%$, $SD = 20.3$). The interaction between Validity and Believability did not reach significance, $F(1,37) = 4.03$, $p = .052$, $\eta_p^2 = .098$. There was no main effect of Pain, $F(2,74) = .37$, $p = .694$, $\eta_p^2 = .010$. All other effects were non-significant, $p$s > .195, $\eta_p^2 < .044$.

A Bayesian version of this ANOVA using Rouder et al.'s [36] default prior width of h = 0.5 revealed that the best-fitting model included Believability, Validity, and Believability × Validity, P(model|data) = .73. There was very strong evidence against the model that additionally included Pain, BF_01 = 32.22, and very strong evidence against the model that additionally included Pain and all interactions with Pain, BF_01= 58.46.

*4.2.3.4. Belief Bias Syllogisms (confidence ratings)*

A 3 (Pain: none, low-intensity, high-intensity) × 2 (Validity: Valid, Invalid) × 2 (Believability: Believable, Unbelievable) ANOVA on syllogism confidence ratings showed a significant main effect of Validity, $F(1,37) = 16.88$, $p < .001$, $\eta_p^2 = .313$, with higher confidence for valid inferences ($M = 2.37$, $SD = 0.39$) than invalid inferences ($M = 2.18$, $SD = 0.43$). All other effects were non-significant, $p$s > .054, $\eta_p^2 < .076$.

A Bayesian ANOVA on syllogisms confidence ratings with Believability, Validity, and Pain Condition as factors and Rouder et al.'s [36] default prior width of h = 0.5 found that the

best-fitting model included only Validity, P(model|data) = .70. There was moderate evidence against the model that additionally included Pain, BF_01 = 5.00, and strong evidence against the model that additionally included Pain and its interaction with validity, BF_01= 13.69.

### 4.3. Experiment 3 Discussion

Experiment 3 showed no effects of pain on syllogisms performance, syllogisms confidence, CRT performance, or Raven's Matrices performance. Our hypothesis that reasoning performance would be poorer in the low intensity pain condition relative to the no pain condition, and better in the high intensity pain condition than low intensity pain condition, was therefore not supported.

## 5. General Discussion

In a series of three experiments we found no clear evidence to support the hypothesis that laboratory-induced pain negatively influences reasoning, in particular the ability to inhibit intuitive but incorrect responses (Cognitive Reflection Test), the ability to distinguish valid and invalid arguments with belief-laden content (the Belief Bias Syllogisms Task), and the ability to distinguish valid and invalid deductions from abstract conditional statements (Conditional Inference Task). We found no evidence to support our general argument that the small but persistent negative effect of pain on attention and executive functioning should lead to negative effects of pain on tests of logical reasoning, since those tasks of reasoning necessarily involve 'Type 2' processes such as working memory updating, attentional switching, and inhibition of automatic responding.

Reporting and interpreting negative findings is always fraught with interpretative danger, although it is a necessary task [25]. There are two broad conclusions one might draw from these findings. First, that our hypotheses are incorrect and that reasoning, as captured in these

tasks, is unaffected by induced pain. Second, that there is an effect of pain on reasoning, but it is not shown using these methods (i.e., we made a Type II error). We discuss each of these possibilities in turn.

If reasoning is not affected by pain, this could be explained in several ways. It may be that some of the component processes involved in reasoning are unaffected by pain and able to compensate for processes which are affected by pain. Or, it may be that participants find these tasks engaging enough to hold their attention away from the pain [45]. Alternatively, it may be that effort on these tasks increases in a linear fashion as pain increases, so that any disruptive effect of pain on task performance is counteracted by increased effort. We suggest that the first two explanations are more plausible than the third. While the first two explanations are consistent with finding disruptive effects of pain on more basic attention tasks that have been studied previously, but null effects on reasoning tasks, the third does not clearly differentiate between different types of tasks. Our motivation for Experiment 3 was that high-intensity pain may prompt participants to put more effort into the task, whereas lower intensity pain may not. In this case we would expect lower-intensity pain to be associated with poorer task performance and higher-intensity pain to be associated with better task performance than low-intensity pain (it could be better, poorer or similar to performance without pain). We did not find any evidence for this hypothesis. If effort did increase linearly with pain intensity then that would be consistent with our null results, but difficult to reconcile with previous studies showing effects of pain on attention tasks.

In order for us to confidently assert that pain does not disrupt reasoning, we need further evidence of non-effects, with independent replication of these experiments using identical tasks and procedures, and conceptual replications using other reasoning tasks and other painful stimuli, with Bayes factors providing support for the null hypothesis. Uncertainty is reduced as negative effects amass.

If there is an effect of pain on reasoning, but we made a Type II error in failing to reject the null hypothesis, then the promising candidates for the source of our error are the primary task(s), the pain challenge, the sample size and the experimental context. Most tests of attention and memory provide indices of error and response time, and we expect in many situations a strategic trade-off between them. Here we focused on error only. It is possible that people compensate with time in order to maximise performance. Indeed, the finding that time pressure has been shown to influence task performance is relevant [12; 13; 22]. Previous studies that have considered both response times and accuracy have not found clear evidence of a trade-off. In some cases, pain is associated with lower accuracies but no significant change to response times [6], while in other cases pain is associated with both lower accuracies and longer response times [4,7], or, consistent with a speed-accuracy trade-off, longer response times and no significant effect on accuracies [4,7]. Given these inconsistencies, we recommend that future research considers the effects of pain on response times as well as on error rates.

Some of our tasks suffered from a lack of room for variance, particularly the CRT and syllogisms confidence ratings in Experiment 3. It is possible that with more items in the CRT, and a wider scale for giving confidence ratings, these analyses would have shown effects of pain. On the other hand, the Bayesian analyses of these measures found moderate evidence against effects of pain.

We used only one pain challenge in these experiments. Although we made sure it was painful, the pain was short-lived with a distinct temporal on- and off-set. It is possible that a sustained pain stimulus, such as a cold-pressor, would be a more complex and effective challenge. Further, most laboratory tasks are *de facto* unthreatening ─ the key function of pain to signal harm is diminished. Needed here, perhaps, are tasks that are a closer analogue to real-world pain [31]. It is possible that naturally-occurring pain would influence reasoning

behaviour in a way that experimental pain did not in these experiments. Recent studies have found that while experimental pain did not change abstract thinking behaviour [1], the intensity of clinical pain did [23]. The benefit of using experimental pain is that we can isolate the effects of pain from the effects of confounds such as other health problems, depression, and pain-related anxiety. Where it is found that clinical pain but not experimental pain disrupts cognition, this may suggest that the cause of the disruption is not pain itself, but something that commonly co-occurs with pain in clinical samples. Our use of exclusively experimental pain limits the generalisability of our findings, but provides an important element of our larger understanding of how pain interacts with cognition.

Our sample sizes were modest, but given that there were no consistent non-significant trends between the three experiments, and that our Bayesian analyses found support for the null hypothesis in many cases, lack of power is unlikely to be the reason we failed to reject the null hypotheses. We have not computed observed power or sensitivity due to issues that have been identified with this approach [24]. Nevertheless, future investigations should be powered to find the smallest effect sizes of interest, and sample sizes and analysis plans should be pre-registered on a service such as the Open Science Framework (www.osf.io) or As Predicted (www.aspredicted.org) [30].

Finally, the participants were students at a University, and may be less susceptible to errors of reasoning, or more primed to be careful in task performance. Recruiting from the general population would be beneficial in this regard.

We are keen to broaden the study of the influence of pain on cognition and behaviour to include macro-investigation of people making real decisions, in addition to the micro-investigations of error rates and response times on laboratory tasks. Despite these negative findings, we believe that investigating the effect of pain on real-world decisions is the right direction of travel for this research. This will inevitably mean that we need to invest in

methods development, in the measurement of key attentional features such as interruption [2], in better human pain models [31] and in sampling from people with pain in both planned and accidental acute pain environments, and in various chronic pain environments. Similarly, we will need to make our tasks more relevant to real world goals, for example attending to goal-relevant information [28; 45] or reasoning about analgesic decisions.

If future work goes on to find support for the hypothesis that pain disrupts higher-level cognitive processes, such as reasoning, it will be important to investigate the role of biopsychosocial factors in this effect, including sex and gender[18; 26; 34], and pain-related anxiety, fear and catastrophizing[44; 47].

In summary, we report three experiments showing no significant effect of pain on reasoning. We are unable to reject the null hypothesis that reasoning is unaffected by pain. We are committed to the publication of null results and recognize that this is increasingly rare in science [16]. However, the major challenge remains one of interpretation. The absence of any effect is theoretically troubling, and before accepting the null hypothesis, we suggest further investigation. Candidate next steps are to broaden the measurement of reasoning, to introduce longer duration pain protocols, and to make task demands motivationally salient.

**Acknowledgments**

This research was funded by an unrestricted grant from Reckitt Benckiser UK Commercial Ltd to CE and EK. The authors would like to thank Joseph Walsh for help with data collection for Experiment 3. CE and EK report grants from Reckitt Benckiser UK Commercial Ltd during the conduct of the study. EK has provided unrelated consultancy services to Reckitt Benckiser UK Commercial Ltd. NA has nothing to disclose. The authors have no other conflicts of interest to declare.

Figure 1. Percentage of valid and invalid items endorsed as a function of their believability in Experiment 1. Error bars represent ±1 standard error of the mean.

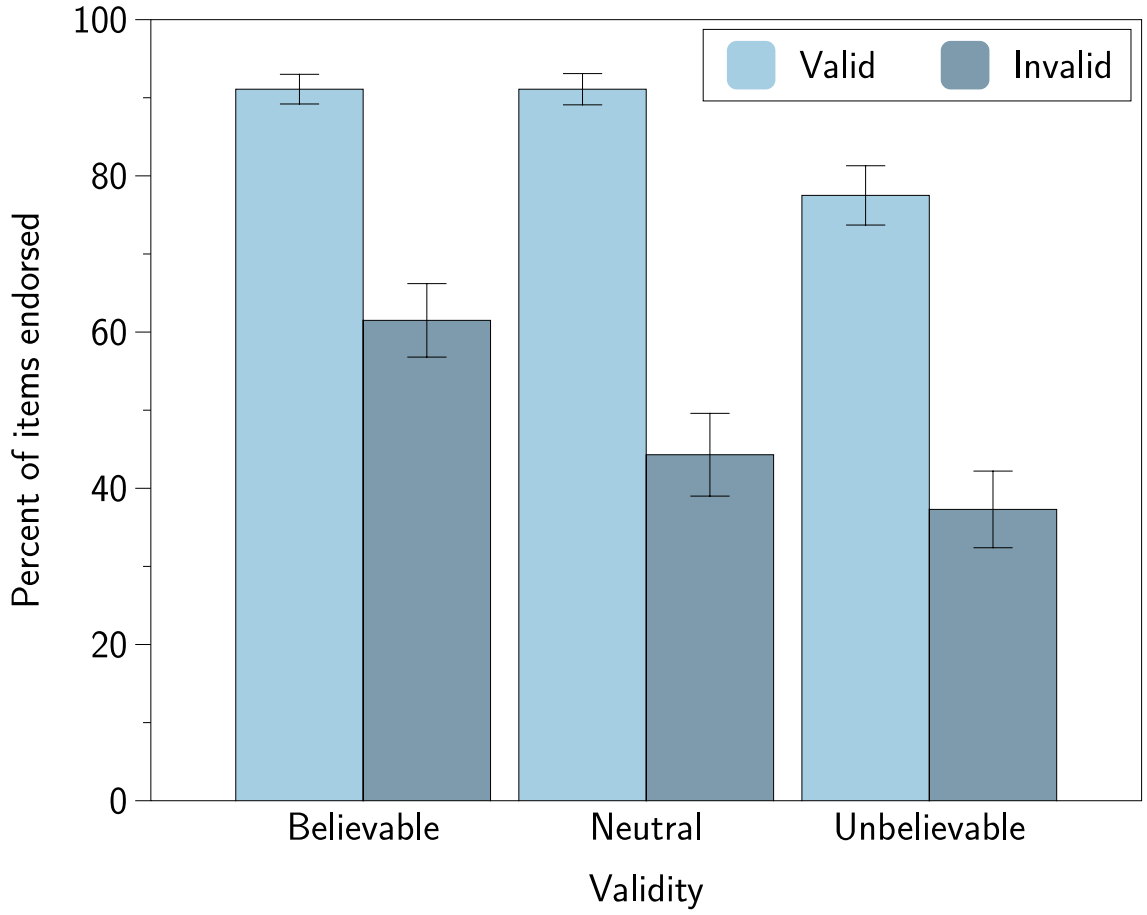
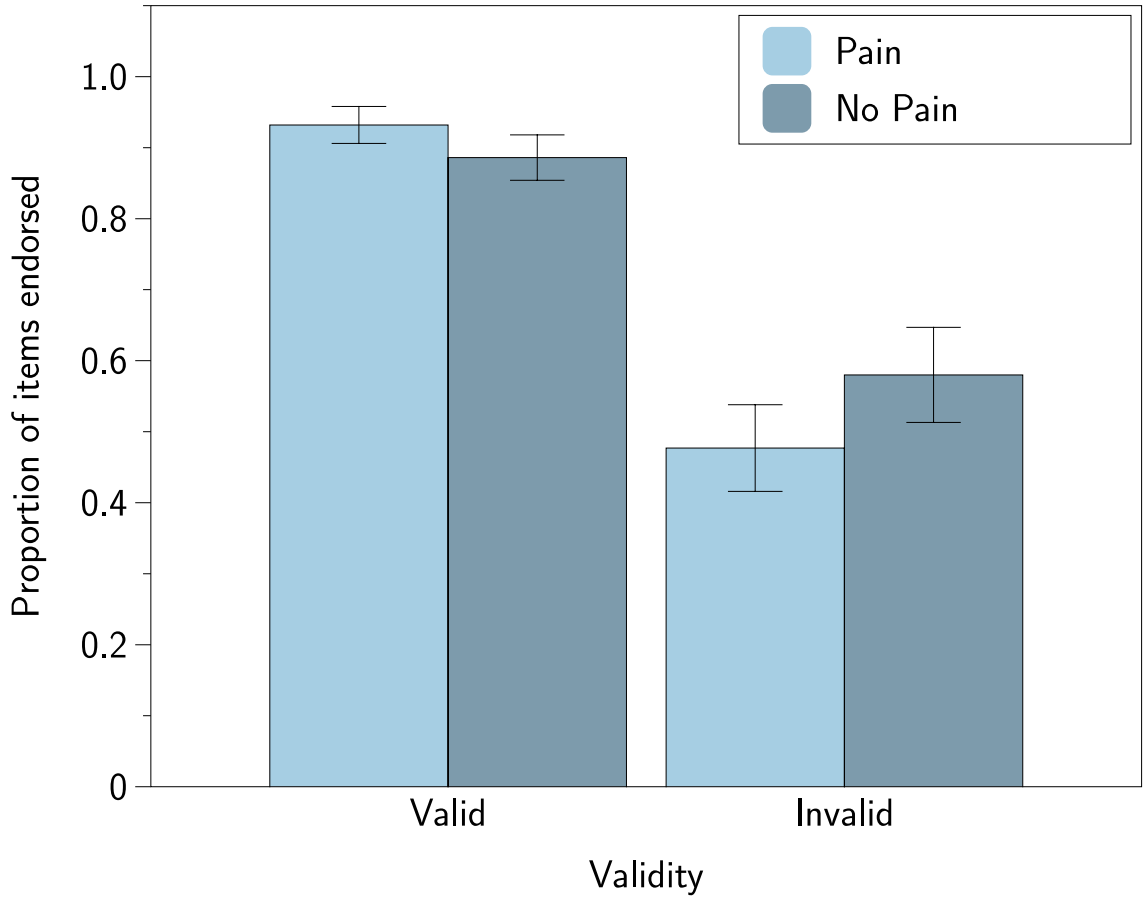


Figure 2. Pain by Validity interaction within the believable syllogisms in Experiment 2. Error bars represent ±1 standard error of the mean.



# References

[1] Agerström J, Gunnarsson H, Stening K. Does physical pain impair abstract thinking? Journal of Cognitive Psychology 2017;29:748-754.

[2] Attridge N, Crombez G, Van Ryckeghem D, Keogh E, Eccleston C. The Experience of Cognitive Intrusion of Pain: scale development and validation. Pain 2015;156:1978-1990.

[3] Attridge N, Doritou M, Inglis M. The development of reasoning skills during compulsory 16 to 18 mathematics education. Research in Mathematics Education 2015;17:20-37.

[4] Attridge N, Eccleston C, Noonan D, Wainwright E, Keogh E. Headache impairs attentional performance: a conceptual replication and extension. The Journal of Pain 2017;18:29-41.

[5] Attridge N, Inglis M. Increasing cognitive inhibition with a difficult prior task: implications for mathematical thinking. ZDM Mathematics Education 2015;47:723-734.

[6] Attridge N, Keogh E, Eccleston C. The effect of pain on task switching: pain reduces accuracy and increases reaction times across multiple switching paradigms. PAIN 2016;157:2179-2193.

[7] Attridge N, Noonan D, Eccleston C, Keogh E. The disruptive effects of pain on n-back task performance in a large general population sample. PAIN 2015;156:1885-1891.

[8] Crombez G, Eccleston C, Baeyens F, Eelen P. Attentional disruption is enhanced by the threat of pain. Behaviour Research Therapy 1998;36:195-204.

[9] Evans J, Clibbens J, Rood B. Bias in conditional inference: implications for mental models and mental logic. The Quarterly Journal of Experimental Psychology 1995;48:644-670.

[10] Evans JSB. In two minds: dual-process accounts of reasoning. Trends in Cognitive Sciences 2003;7:454-459.

[11] Evans JSB, Barston JL, Pollard P. On the conflict between logic and belief in syllogistic reasoning. Memory & Cognition 1983;11:295-306.

[12] Evans JSB, Curtis-Holmes J. Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. Thinking & Reasoning 2005;11:382-389.

[13] Evans JSB, Handley SJ, Bacon AM. Reasoning under time pressure: A study of causal conditional inference. Experimental Psychology 2009;56:77-83.

[14] Evans JSB, Newstead S, Allen J, Pollard P. Debiasing by instruction: The case of belief bias. European Journal of Cognitive Psychology 1994;6:263-285.

[15] Evans JSB, Stanovich KE. Dual-process theories of higher cognition: Advancing the debate. Perspectives on Psychological Science 2013;8:223-241.

[16] Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics 2011;90:891-904.

[17] Field A. Discovering statistics using IBM SPSS statistics. 4th Ed. 2013; London: Sage.

[18] Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley Iii JL. Sex, Gender, and Pain: A Review of Recent Clinical and Experimental Findings. The Journal of Pain 2009;10:447-485.

[19] Frederick S. Cognitive reflection and decision making. The Journal of Economic Perspectives 2005;19:25-42.

[20] Gilhooly K, Logie R, Wetherick N, Wynn V. Working memory and strategies in syllogistic-reasoning tasks. Memory & Cognition 1993;21:115-124.

[21] Gilhooly K, Logie RH, Wynn V. Syllogistic reasoning tasks, working memory, and skill. European Journal of Cognitive Psychology 1999;11:473-498.

[22] Gillard E, Van Dooren W, Schaeken W, Verschaffel L. Proportional reasoning as a heuristic-based process: Time constraint and dual task considerations. Experimental Psychology 2009;56:92-99.

[23] Gunnarsson H, Agerstöm J. Clinical pain, abstraction, and self-control: being in pain makes it harder to see the forest for the trees and is associated with lower self-control. Journal of Pain Research 2018;11:1105-1114.

[24] Hoenig JM, Heisey DM. The abuse of power: The pervasive fallacy of power calculations for data analysis. The American Statistician 2001;55:19-24.

[25] Ioannidis JP, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends in Cognitive Sciences 2014;18:235-241.

[26] Keogh E. Sex and gender differences in pain: a selective review of biological and psychosocial factors. The Journal of Men's Health & Gender 2006;3:236-243.

[27] Keogh E, Cavill R, Moore DJ, Eccleston C. The effects of menstrual-related pain on attentional interference. PAIN 2014;155:821-827.

[28] Keogh E, Moore DJ, Duggan GB, Payne SJ, Eccleston C. The disruptive effects of pain on complex cognitive performance and executive control. PloS One 2013;8(12):e83272.

[29] Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology 2013;4:863.

[30] Munafò, MR, Nosek, BA, Bishop, DV, Button, KS, Chambers, CD, du Sert, NP, ... & Ioannidis, JP. A manifesto for reproducible science. Nature Human Behaviour 2017;1:0021.

[31] Moore DJ, Keogh E, Crombez G, Eccleston C. Methods for studying naturally occurring human pain and their analogues. PAIN 2013;154:190-199.

[32] Moore DJ, Keogh E, Eccleston C. The interruptive effect of pain on attention. The Quarterly Journal of Experimental Psychology 2012;65:565-586.

[33] Moore DJ, Keogh E, Eccleston C. Headache impairs attentional performance. PAIN 2013;154:1840-1845.

[34] Paller CJ, Campbell CM, Edwards RR, Dobs AS. Sex-Based Differences in Pain Perception and Treatment. Pain Medicine 2009;10:289-299.

[35] Raven J, Raven J, Court J. Manual for Raven's progressive matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment, 2000.

[36] Rouder JN, Morey RD, Verhagen J, Swagman AR, Wagenmakers EJ. Bayesian analysis of factorial designs. Psychological Methods 2017;22:304-321.

[37] Sá WC, West RF, Stanovich KE. The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. Journal of Educational Psychology 1999;91:497.

[38] Stanovich KE. The robot's rebellion: Finding meaning in the age of Darwin: University of Chicago Press, 2004.

[39] Stanovich KE, Cunningham AE. Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. Memory & Cognition 1992;20:51-68.

[40] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory & Cognition 2011;39:1275-1289.

[41] Toplak ME, West RF, Stanovich KE. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking & Reasoning 2014;20:147-168.

[42] Trippas D, Handley SJ, Verde MF. The SDT model of belief bias: complexity, time, and cognitive ability mediate the effects of believability. Journal of Experimental Psychology: Learning, Memory, and Cognition 2013;39:1393-1402.

[43] Trippas D, Pennycook G, Verde MF, Handley SJ. Better but still biased: Analytic cognitive style and belief bias. Thinking & Reasoning 2015;21:431-445.

[44] Van Damme S, Crombez G, Eccleston C. Disengagement from pain: the role of catastrophic thinking about pain. PAIN 2004;107:70-76.

[45] Van Damme S, Legrain V, Vogt J, Crombez G. Keeping pain in mind: A motivational account of attention to pain. Neuroscience & Biobehavioral Reviews 2010;34:204-213.

[46] Van Ryckeghem DM, Crombez G, Eccleston C, Liefooghe B, Van Damme S. The interruptive effect of pain in a multitask environment: an experimental investigation. The Journal of Pain 2012;13:131-138.

[47] Vlaeyen JWS, Linton SJ. Fear-avoidance model of chronic musculoskeletal pain: 12 years on. PAIN 2012;153:1144-1147.